Measuring 'Virtue' in Medicine

Kotzee, Hendrik; Ignatowicz, Agnieszka

DOI:
10.1007/s11019-015-9653-6

License:
Other (please specify with Rights Statement)

Document Version
Peer reviewed version

Citation for published version (Harvard):
Kotzee, H & Ignatowicz, A 2015, 'Measuring 'Virtue' in Medicine', *Medicine, Health Care and Philosophy*, vol. 19, no. 2, pp. 149–161. https://doi.org/10.1007/s11019-015-9653-6

Measuring 'Virtue' in Medicine

1. Introduction

A growing body of work stresses the importance of virtue concepts to medical ethics (Pellegrino and Thomasma, 1993; Oakley and Cocking, 2001; Toon, 2014). According to virtue theoretic approaches, questions of morality or right action in medicine are settled in terms of the moral character of the doctor performing the action. What is right or moral to do in medicine is what the virtuous doctor would do in a specific situation. A different theme in medical ethics is how to understand doctors' ethical decision-making empirically. Studies of doctors' ethics have been conducted using a range of methods – from psychometrically validated measurement instruments to observations, peer-reports, interviews and records of disciplinary proceedings (Baldwin and Self, 2005; Bebeau, 2006; Jha *et al.*, 2007; Wilkinson *et al.*, 2009). However, most measurement instruments, assessments or tests in the area are *cognitive* and *not* virtue-based. Such instruments focus on doctors' skills in reasoning about and articulating principles for ethical action and not on their moral character as professionals. In this respect, developments in theoretical and empirical approaches to medical ethics are out of alignment: while the literature on medical ethics increasingly stresses the importance of virtue or moral character in a doctor, currently available empirical instruments focus on measuring doctors' moral reasoning. In this paper, we argue that virtue-based assessment tools are sorely needed if virtue approaches to medical ethics are to make an impact in fields like medical education and assessment.

The paper proceeds as follows. In section 2, we sketch why work on the theory of medical ethics needs to be accompanied by empirical research if it is to make an impact in fields like medical education and regulation. In sections 3 and 4, we survey the literature on moral development in psychology and on the empirical study of ethics in medicine and conclude that most work on the moral development of doctors still departs from a cognitive perspective. In section 5, we sketch out the prospects (and also the problems) that exist for designing virtue-based measures of doctors' moral development. We find that theoretical interest in virtue or character in medical ethics is not yet matched by suitable tools to assess the development of virtuous character traits in doctors. In section 6 we discuss the prospects for future research in the area. Section 7 concludes.

2. Why measure virtue?

What can be termed the 'virtue turn in medical ethics' can be traced to the work of authors such as Pellegrino and Thomasma (1993) who express dissatisfaction with rules or principles-based ethics ('deontology') and offer, in its place, a conception of medical ethics that focuses on the virtues of the good doctor. In asking what makes a good or professional doctor, virtue-based accounts of medical ethics turn their attention away from features of the action being performed (whether, say, a treatment satisfies principles like being beneficial, just, respecting autonomy, etc.) to the moral character of the practitioner making the decision or performing the treatment (to whether that *person* is honest, kind, just, etc. in how they act). There is growing support for a virtue-based approach to medical ethics (Coulehan, 2005; Bryan and Babelay, 2009) and today, medical ethics finds itself faced by a choice as

to the purpose of education, regulation and monitoring in the field of medical ethics (Eckles *et al.* 2005: 1145): should such programmes be concerned with:

(i) the ability of medical practitioners to reason well in terms of principles of medical ethics

or should it be concerned that

(ii) medical practitioners be virtuous doctors?[1]

This brings us to the matter of assessment – the topic of this paper. At first glance, the idea that one may measure or assess the character of a doctor empirically seems queer. Talk of measuring and policing doctors' moral character may also seem politically objectionable. What business is it of professional and regulatory bodies what a doctor's moral character is like?, one may ask. Moreover, who is to determine what constitutes acceptable moral character for a doctor and whether a particular doctor meets this character test?

Taking a virtue-perspective to medical ethics brings one to confront exactly these questions, for if right action in medicine is settled in terms of the character of the doctor that performs action, it will be necessary to *tell* whether a doctor is virtuous in order to determine whether they practice morally or well. If we take a virtue-perspective on medical ethics, then education in the field of medical ethics will have as its purpose the formation of virtuous doctors. Furthermore, ongoing regulation and monitoring in the field of medical ethics will be concerned to ensure that only doctors of acceptable moral character continue to practice and to make available remediation to or, ultimately, bar from practice those doctors whose moral character (*qua* doctor) is not acceptable. Doing either of these things – educating virtuous doctors or ensuring good character amongst doctors – requires that we be able to assess doctors' character empirically to an acceptable standard. This is so for three reasons:

> *Firstly*, without being able to understand professionals' character as it is, it will be hard to know what it is about medical students' and young doctors' character that needs to be shaped through appropriate education in medical ethics.
>
> *Secondly,* without being able to assess how a student or doctor's character changes or has changed, it will be impossible to know whether any educational intervention designed to shape their character has been effective.
>
> *Thirdly*, without being able to assess practicing doctors' character, it will be impossible to identify the minority (one hopes) of doctors who deserve either remediation or sanction in terms of their moral character.

If they are to make an impact in fields like the education and regulation of doctors, then, virtue-based approaches to medical ethics need to develop valid, reliable and fair means to assess doctors' moral character.

[1] Eckles *et al.* (2005) call this the 'virtue/skill dichotomy'.

Empirical resources to this purpose, however, are scant. Other authors have reported that it is not always clear how professional ethics in medicine should be taught and how it should be assessed is a matter that is even less clear.[2] Moreover, most of the resources that are available for understanding doctors' ethics proceed from an understanding of medical ethics that is rules- or principles-based and – therefore – conceives of moral psychology as a cognitive matter. No agreed method to understand the moral character of professionals exists, meaning that medical ethicists who adopt a virtue ethics perspective are flying blind in their efforts to shape the character of medical students and to regulate the character of practising doctors.[3]

In explaining the measurement challenge for virtue ethics in medicine, we will first give an overview of past attempts to understand morality in psychology broadly speaking before confining our attention to understanding virtue or moral character in medicine.

3. Measuring morality in psychology and medicine

To anyone working in moral psychology, - education or - philosophy, Lawrence Kohlberg's (1981; 1984) work on moral development will be deeply familiar. Kohlberg posits that moral development from childhood into adulthood unfolds through an invariant sequence of modes of moral thinking, from thinking of right action in terms of self-interest (pre-conventional thinking) to thinking of right action in terms of what is socially desirable (conventional thinking) to thinking of right action in terms of a set of moral principles that one can rationalise as universally valid (post-conventional thinking). The last form of moral thinking is clearly *rules*- or *principles*-based and, for Kohlberg, it constitutes the highest form of moral thinking. Kohlberg's Moral Judgement Interview (MJI) was designed as a test of development of moral *cognition*. It aims to capture the extent to which individuals have progressed from pre-conventional through conventional to post-conventional thinking and still constitutes the purest expression of measuring moral development according to the Kohlbergian approach.[4]

Kohlberg's research programme came to be very influential in the study of professional ethics. Especially Kohlberg's colleague, James Rest's (1999) four component model of moral development played a crucial role in the field. According to Rest's model, moral action results from a combination of

- moral sensitivity
- moral motivation
- moral judgement

[2] See Calman (1987) Cowley (2005b) and Mattick and Bligh (2006).

[3] *Some* virtue ethicists may hold that virtue is not the sort of thing that one can test for using a psychometric test, but that whether another person is virtuous is a matter of judgement. This is an interesting matter that cannot be taken up fully here. Suffice it to say that, were it true that virtue is not susceptible to empirical testing, this would leave virtue approaches at a permanent disadvantage *vis-à-vis* rules-/principles-based approaches, as such approaches *are* thought to be amenable to testing. This may not trouble all virtue ethicists; however it behoves virtue ethicists not to give up on the possibility of empirical testing of virtue too readily.

[4] In this paper, we often write of rules- or principles-based approaches to ethics and cognitive approaches to moral psychology in the same breath. The one naturally invites the other.

- moral character

The emphasis in this tradition is on studying *moral schemas* – typical, acquired and routine ways that moral subjects have of responding to moral problems. In order to study the moral schema that a person brings to a task, researchers in the four-component tradition devised a test of moral development, the Defining Issues Test (DIT). As originally conceived, the DIT was designed to measure only moral judgement schemas and not to measure the other three of Rest's four components. A second version of the DIT (the DIT-2) was later developed and the DIT(2) remains the most extensively used instrument to study moral reasoning.

Kohlbergian moral psychology has had a profound effect on the empirical study of medical ethics and most psychometric measures of professional ethics in medicine are Kohlbergian in their design or inspiration. Broadly speaking, existing measures can be grouped in the following classes.

Measures of General Moral Reasoning

Self and Baldwin (1994), Baldwin and Self (2005) and Bebeau (2002, 2006) survey the literature on ethics assessments in medical education[5] and conclude that measures of general moral reasoning in the Kohlbergian tradition are the most influential in shaping thinking about how to assess doctors' ethics empirically. Such measures put completely general moral dilemmas (i.e. dilemmas not confined to the medical field) to respondents and evaluate their ability to engage in moral reasoning regarding these. Self and Baldwin (1994) and Baldwin and Self (2005) review a total of 40 studies on moral reasoning in medicine. Of those studies 25 used the DIT, 5 the MJI and 3 another the SRM. Bebeau (2002) reviews 33 studies (themselves involving 6,600 participants) of the use of such tests in professional education of which 22 concern medicine. In these studies the DIT is by far the most used instrument. More recent work since Bebeau's 2006 review article (e.g. Patenaude, Niyonsenga and Fafard, 2003; Helkama et al., 2003 and Hren *et al.*, 2011) does not belie this general trend: studies of doctors' ethics mostly employ general measures of moral reasoning (and most of these use the DIT).

Kohlbergian Profession-Specific Measures of Moral Reasoning

Next to measures of general moral reasoning, some profession-specific measures of moral reasoning (containing dilemmas only relevant to one profession) exist. The best example is the 'Dental Ethical Reasoning and Judgement Test' (DERJT). The DERJT attempts to measure the ability to apply profession-specific ethical concepts – like 'confidentiality', 'conflict of interest', 'consent', etc. – to common ethical cases which dentists may face in their job (Bebeau and Thoma 1999). The theoretical underpinnings of the DERJT lie with the neo- or post-Kohlbergian Four Component Model of ethical decision-making sketched earlier. Like the more generic DIT whose theoretical framework it shares, it operationalizes only the second component entering ethical behaviour, namely 'Ethical Judgment/Reasoning'. Today, the DERJT is widely used for purposes of selecting candidates to dental schools in the US. (Lantz, Bebeau et al. 2011).

[5] …and professional education more broadly, in the case of the papers by Bebeau.

Given the success of the DERJT in the field of dentistry, there have been attempts to design a similar measure for medicine. Two intermediate concept measures for medicine exist: the Medical Ethical Reasoning and Judgement Test (MERJT) (Caldicott, Faber-Langendoen, Bebeau and Thoma, 2008/2010) and the Medical Intermediate Concept Measure of Ethical Reasoning (MD-ICM). (Pinijphon, 2009)

Other Kohlbergian Profession-Specific Measures

While most measures in the Kohlbergian and post-Kohlbergian traditions focus on moral reasoning, some tests also exist of moral sensitivity. Bebeau, for instance, has developed two measures focussing on separate components of Rest's four component model: the Dental Ethical Sensitivity Test (DEST) was developed as a test of ethical sensitivity in dentistry and the Professional Role Inventory (PROI) was developed as a test of the professional's identification with their role. Apart from DEST and PROI, other attempts have been made to assess ethical sensitivity. Hébert *et al.* (1990, 1992) developed a questionnaire to measure medical students' and professionals' ethical sensitivity, which Akabayashi and colleagues (2004) later translated, condensed and adapted to the Japanese clinical setting (currently referred to as the Problem Identification Test or PIT).

Non-Kohlbergian Profession Specific Measures of Moral Reasoning

Lohfeld *et al.* (2012) describe the development of an instrument to serve as an alternative to broadly Kohlbergian tests of moral development like the MJI and DIT – the Ethics in Health Care Questionnaire version 2 (EHCQ-2). Based on the Ethics in Health Care Questionnaire (EHCQ) (Kipnes and Gerhard, 1995) designed for use with nurses, the new version of the EHCQ-2 has been trialled with 65 medical students in Scotland and Canada. The authors present the EHCQ-2 as an *alternative* to Kohlbergian measures (such as the MJI and DIT) of moral development. The reasons that they put forward for designing such an alternative are that there exist theoretical problems for Kohlberg's model of stage-wise moral development and that these

> '…instruments do not explicitly measure people's ethical values or their ability to use them in real-life situations, but instead use test scores to classify people's progress along an abstract and controversial moral development scale…' (Lohfeld *et al.* 2012: 636).

However, in *design*, the EHCQ-2 is not all that dissimilar to Kohlbergian instruments. It uses the Kohlbergian model of setting the respondent a dilemma and asking what course of action the respondent would take. The respondent is then asked to provide a free response reason as to why they would take this course of action. It is highly plausible that, in asking the respondent for a reason why they would act as they say they would, the EHCQ-2 still primarily tests the respondent's *moral reasoning*. Certainly, no effort is made to elicit something more immediate, like a purely emotional response to the dilemma or to measure one of the other three components of moral decision-making besides moral reasoning (i.e. moral sensitivity, motivation and character).

A further new test that deserves mention is the Ethical Reasoning Inventory (Tsai *et al.*, 2009). Like Lohfeld *et al.* (2012), Tsai and colleagues bemoan the lack of medicine specific measures and the reliance on general measures like the DIT and

MJI. The ERI presents test takers with ethical vignettes in an interview setting. Test takers are instructed to 'think aloud' while solving the moral problems presented in the ethical vignettes.

Assessments that focus on performance

A number of assessments aim to measure doctors' performance in ethical situations more directly (rather than asking what they would do in hypothetical dilemmas such as those contained in the paper-based and interviewing-based studies above). Smith *et al.* (1994) describes a performance-based assessment of ability in moral reasoning and ethical judgement using a structured patient encounter. Similarly, Ramakrishna, Sriharan and Scolnik (2014) developed and piloted a 31 item tool to evaluate professionalism, communication and collaboration amongst medical students. The authors collected data through faculty's assessment of students, peer assessment and self-assessment. They concluded that the tool was sensitive to change as students improved in each competency, the greatest change being in collaboration. The qualitative data also seemed to support the observed changes.

Other assessment instruments and non-validated approaches

A large number of studies have been conducted of the concept (related to ethics) of medical 'professionalism'. In their systematic review of studies of the assessment and development of professionalism in medicine, Jha *et al.* point out that the definition of 'professionalism' in medicine is somewhat vague:

> 'much of the literature defines professionalism in vague terms - altruism, humanism, excellence - few studies have operationalized medical professionalism' (Jha *et al.*, 2006: 1027).

Due in no small part to the difficulties involved in operationalising the concept of professionalism, Jha *et al.*'s systematic review of studies from the USA and the UK finds that few are able to prove a link between teaching interventions and a change in attitudes towards professionalism. (Jha, Bekker, Duffy, & Roberts, 2007)[6] Jha *et al.* (2007: 825) found 44 studies reporting the use of scales to measure attitudes towards various aspects of professionalism. Aspects measured were attitudes to the patient–physician relationship, to values, integrity, cultural issues, social issues, other professions, general professionalism, humanities, ethics, crossing boundaries and moral issues. Wilkinson *et al.* (2009) identify nine broad forms (or 'clusters') of types of assessment used to measure professionalism: observed clinical encounters, collated views of coworkers, records of incidents of unprofessionalism, critical incident reports, simulations, paper-based tests, patients' opinions, global views of supervisor, and self-administered rating scales. Wilkinson *et al.* do not mention character or virtue specifically as a component of professionalism. Furthermore, in their discussion of 'paper-based tests' of professionalism, the DIT again surfaces as one of the most frequently used tests. (Wilkinson, *et al.* 2009: 554) Neither of these two reviews of assessments of professionalism (by Jha *et al.* and Wilkinson *et al.*), then, reveal a great concern with measuring character or virtue as part of measuring professionalism.

[6] Some of the problems involved in measuring professionalism or ethical awareness/behaviour are already set out in Eccles, 2005.

4. From Moral Cognition to Moral Virtue

Above, we saw that empirical research into doctors' moral decision-making is dominated by research regarding doctors' moral *reasoning.* A number of criticisms of the cognitive research programme in moral psychology exist.

i. Firstly, the link between moral reasoning and moral *behaviour* is weaker than the cognitivists suppose. According to the cognitive programme, it is conscious reasoning about what to do that drives moral action. As Kohlberg himself put it, '[h]e who knows the good chooses the good' (1981: 189). Working at the same time as Kohlberg, Blasi (1980), however, raised serious doubts about whether or not this is true. Blasi conducted a meta-analysis of studies on the link between moral cognition and moral action and found few correlations between Kohlbergian stages of moral reasoning and actual moral behaviour. Blasi concluded that moral reasoning is not the only determinant of moral behaviour (Sanderse, 2012: 49); pertinently, moral emotion is another important factor (see below).

ii. Secondly, the cognitivists' emphasis on moral development occurring in stages has been subject to some criticism. Recall that, for Kohlberg, moral development occurs according to a predictable trajectory from pre-conventional moral thinking through conventional moral thinking to post-conventional moral thinking. The highest stage of moral development – post-conventional moral thinking – is characterised by taking a Kantian approach to ethics, that is, to seek to fit one's own moral behaviour to universalizable moral principles (and expecting the same of others). Criticism has been voiced against (a) the idea that moral development takes place for all people and for all cultures in exactly this same stage-like way *and* (b) against the idea that Kantian moral thinking constitutes the highest form of moral thinking. Some studies have shown that there can be inconsistency in individual moral development in that the same individual may appear to be in different stages of moral development, depending on the kind of moral dilemma they are confronted with (see, e.g. Carpendale, 2000). Other studies have shown that there is some cultural variability to moral development. (Harkness, Edwards and Super, 1981) Moreover, the basic idea that the highest form of moral development consists in taking a Kantian approach to ethics is criticised on philosophical grounds. The debate in ethics between deontologism, consequentialism and virtue ethics is by no means settled and each camp contains, in its midst, highly sophisticated moral thinkers. To hold that the Kantian thinking is the most sophisticated is contradicted by this very simple observation regarding the variety of styles of ethical thinking one finds amongst thoughtful people.

iii. Lastly, it is held that the cognitive research programme underestimates the importance of moral emotions or moral feelings next to moral reasoning as determinants of moral action. The 'intuitionist' movement in moral psychology (associated with Haidt and others), stresses that it may not be moral reasoning, but, instead, moral *emotion* that drives moral action. Haidt (2001) reports on studies showing that moral decisions are much more automatic and quick than previously assumed and that moral reasoning is not the well-spring of moral action but represents no more than *post-hoc* rationalisation of a position that an agent has already come to emotionally. As Haidt holds, 'intuition comes first, strategic reasoning second' (2012: 1). Based on these insights Haidt and others have

formulated what is called moral foundations theory (MFT) as an alternative to Kohlbergian theories of moral development and have developed a questionnaire, the Moral Foundations Questionnaire (MFQ) to assess respondents' moral orientation.[7]

Kristjansson (2015: chapter 3) questions whether MFT should be seen as a virtue-approach to the empirical study of morality; in several respects, he holds, it is not. Whereas MFT stresses the emotional well-springs of moral action (and Kohlbergian scholars the rational), according to virtue theorists moral action results from the *alignment of* the emotional and the rational. For the virtue theorist, the moral person is one who experiences the right emotions in a moral situation, but also rationally knows what is the right thing to do. In the virtuous person, these emotions and this knowledge of the good have been built up over time and have become ingrained as part of the person's character to such an extent that they can do the right thing fluently or automatically in a practical moral context. In section 5, we will explain what this implies for the empirical study of virtue; however, the important point for now is that virtue comprises both rational *and* emotional elements that are elicited together in real moral situations. This means that neither Kohlbergian- nor MFT-thinking (or the psychometric research programmes associated with them) fully captures what it is to be virtuous.

What are the distinguishing features of a virtue-based approach to the psychology of morality?

For Aristotle, a virtue is a trait of a person's character (*hexeis*); it is a developed, but (once developed) stable trait that influences the way a person acts from a moral point of view. Some examples of virtues that one would expect doctors to have are honesty, courage, fairness, care, fastidiousness, perseverance, etc. According to an Aristotelian approach, each character trait of this sort consists of a different set of developed tendencies that a person has to do the following things:

- to recognise or perceive moral situations correctly (to be sensitive to what is at stake in a situation)
- to respond emotionally to that situation in the right way (this may include being dispassionate in the right circumstances),
- to think well about what to do in the situation (either to know how to act or to reason appropriately about how to act)
- and to be motivated strongly enough to carry the right action through.[8]

Virtuous action consists in all of these elements operating *in alignment* in a specific situation.

A number of basic problems exist for attempting to study how virtuous a person is scientifically. Firstly, there is the very complexity of the construct of virtue. (Fowers, 2014) In order to determine whether a person is virtuous, one would have to discover

[7] For Haidt, people have different moral orientations or priorities that reflect whether they see a moral problem as one concerning (1) care/harm, (2) fairness/cheating, (3) loyalty/betrayal, (4) authority/suberversion, (5) sanctity/degradation or (6) liberty/oppression. See Haidt (2012).

[8] See Fowers (2005 and 2014) and Kristjansson (2015), for discussion of the psychology of virtue. In a similar vein, Kinghorn (2010) explains a number of these problems from a specifically medical perspective.

whether they see a moral situation aright, whether they show the right moral emotion, whether they reason correctly about the situation, whether they are motivated to do the right thing and whether they do the right thing – all relative to the situation they find themselves in. Measuring all these facets of a person's action in concert in the context of a specific situation is hard. A second problem is that virtue is (to some extent) relative to the individual. Aristotle, for instance, holds that what counts as virtuous behaviour differs from person to person. In Aristotle's example, moderation in eating and drinking is a virtue, but how much to eat or drink is not an absolute matter – a professional wrestler, for instance, needs to eat much more than an ordinary person (NE1106a29 – b4). Similarly with many other virtues. A third problem is that – at least on an Aristotelian view – the virtues are tied up with one another (the principle of the 'unity of the virtues'). (*NE*, 1145a, 1 – 2) If the principle were true, it would imply that assessing for virtue would have to be done on the level of looking for all of the moral virtues together rather than one-by-one. This would be very hard to achieve with standardised psychometric instruments that, typically, aim to test for one psychological construct at a time. A fourth problem is that one cannot be 'fully virtuous' in the moral sense of the word without also displaying the intellectual virtue of *phronesis* (or wisdom or good judgement); nor can one possess *phronesis* without the virtues of good character (NE1144b30–32). In fact, Hursthouse (1999, 2012) explains why this means that moral knowledge is not codifiable in the form of rules at all.[9] Rather than knowing moral rules, the wise moral judge possesses the ability to apply the virtues correctly to real situations on a case-by-case basis. (Hursthouse, 2011: 46) It follows that tests of moral virtue, then, would have to include tests of practical wisdom or phronesis (and *vice versa*).

Given these problems a natural question is whether it is possible to study virtue scientifically at all. One possibility that must be borne in mind is that there simply *are* no stable and measurable character traits that influence the way a person acts from a moral point of view. According to Harman (1999) and Doris (2002) the best interpretation of some results from social psychology (e.g. the 'helping for a dime' experiment, or the 'good samaritan' experiment) is that moral action is situation-dependent to a surprising degree. Because situational stimuli like finding a dime by chance or being in a hurry can make such a big difference to how people act, situationists like Doris and Harman hold that the virtues that we think determine moral action simply do not exist.

The situationist challenge to virtue ethics has received considerable attention in the literature on virtue ethics. Sreenivasan (2002), for instance, holds that the empirical results that Doris and Harman rely on do *not* show that the virtues do not exist, because the cases in which a person can be induced by situational influences to act well or badly are not 'paradigm cases' of the virtues in action. That some subjects may fail to help someone in distress on occasion is still potentially consistent with that person being generally helpful (or having the virtue of kindness). (Sreenivasan, 2002: 60) Kamtekar holds that the experiments that Doris and Harman rely on are only designed to test for the performance or non-performance of certain actions that are stereotypically related to certain virtues in isolation. Instead, for virtue ethicists, virtue exists in demonstrating the right feeling, judgement and action in a particular

[9] Indeed, Aristotle himself cautioned that the study of virtue can be no more exact than its subject matter (i.e. virtue itself) allows for (*NE,* 1094b and 1095a).

situation (Kamtekar, 2004: 477) and Doris and Harman do not give enough attention to *why* the subjects act as they do in the studies cited. In a similar vein, Snow (2010) holds that what evidence we do have regarding the motivation and reasoning of subjects who are induced to act contrary to character by situational forces often shows quite virtuous tendencies (such as having to struggle to overcome the inclination to do the virtuous thing, showing guilt or remorse at not being able to act as they feel they should, etc.) Recently, Jayawickreme *et al.* (2014) have argued, on the basis of empirical evidence gathered via experience sampling methodology, that, if one looks not at isolated or once-off cases of demonstrating a particular behaviour or not, but tracks the same individual's behaviour in similar situations across a period of time, then one *can* establish stable differences between people in the degree to which they exhibit patterns of behaviour. Jayawickreme *et al.* hold that this refutes the situationist criticism of virtue ethics *empirically*.[10]

While it is not possible to settle this issue conclusively here, it appears that there are at least enough indications of the possibility of a psychology of virtue[11] that medical educators and ethicists interested in doctors' virtues and vices may take note of attempts that have been made to study virtue in medicine empirically. In what remains of this paper, we briefly survey attempts in psychology to measure virtue in general and attempts to measure some specific virtues relevant to the practice of medicine – the virtues of wisdom and empathy. We also discuss some work currently in progress.

5. Measuring Virtue in Medicine

Measuring virtue in general

Some studies (e.g. Cawley *et al.*, 2000) adapt personality type scales to the task of studying how virtuous a person is. Cawley *et al.* constructed a virtue scale containing 140 items testing for 140 distinct character traits or virtues. Having collected data about the extent to which people regard themselves as holding these character traits, Cawley *et al.* found that the 140 character traits factor into four main 'virtues': empathy, order, resourcefulness and serenity. A more developed account along similar lines is found in Peterson and Seligman. (2004) Peterson and Seligman distinguish six virtues and 24 'character strengths'. Peterson and Seligman derive their virtues classification from a study of philosophical and religious texts in the Western, Indian and Chinese traditions. (2004: 33 – 52) According to Peterson and Seligman, the 'high six' virtues that can be identified in all of these traditions are: courage, justice, humanity, temperance, transcendence and wisdom. The Values in Action Inventory of Strengths (VIA-IS) associated with the Peterson and Seligman research programme is the most widely used measure in the empirical study of virtue.

Full empirical support for the Peterson and Seligman classification is, however, lacking. Noftle *et al*. (2011), for instance, conclude that the virtues Peterson and Seligman identify do not have the hierarchical structure (of virtues composed of strengths) that Peterson and Seligman think and other studies suggest that, rather

[10] See section 6, below, for more on Jayawickreme *et al.*'s study.
[11] See Curren and Kotzee (2014) for a discussion of the prospects for measuring virtue in moral psychology.

than six factors, Peterson and Seligman's 24 virtues really factor more properly into five (or fewer) factors.[12]

A further problem with the VIA-IS, derives from its nature as a self-report test. Self-report measures are susceptible to both conscious over-reporting of the extent to which one holds a certain character trait and, even, to non-conscious self-deception. Thus, the VIA family of instruments display very high ceiling effects – respondents typically rate themselves very highly on scales purporting to measure how virtuous they are.

We found one study in the field of medical professionalism using the VIA. Schulz *et al.* (2013) report on the development of a virtue-based training programme in leadership and professionalism (Surgical Training and Education in Promoting Professionalism or STEPP) in the Division of Otolaryngology-Head and Neck Surgery (OHNS) at Duke University. Piloted with 9 residents in Head and Neck Surgery in 2011, the STEPP programme consisted of training in five core leadership qualities and offered examples for application of these qualities during resident training. (Schulz, *et al.*, 2013: 2) The programme was evaluated in 2012 and researchers found (*inter alia*) that residents' understanding of the OHNS core values – as measured by the ability to list the five core values – improved from pre-programme to post-programme. The VIA was administered to participants and Schulz *et al.* found a correlation (0.85) between participants' reports on the extent to which others perceive them to be leaders and their VIA score for leadership. Out of the 24 VIA strengths, OHNS residents' mean scores for the strength of curiosity was highest, followed by honesty and judgement.

Arthur *et al.* (2015) conducted a mixed methods study of their own design on character and virtue in medical education and medical practice in the UK. The study sought to understand how medical students and practicing doctors acquire, develop and sustain their professional virtues and asked:

- Which virtues are particularly valued by medical students and doctors?
- How do these virtues shape decisions made in typical moral dilemma situations in medicine?
- How does the environment in students study and doctors practise enable or constrain the exhibition of these virtues?

Arthur *et al.* surveyed 124 undergraduate students, 167 graduating students and 277 experienced professionals and interviewed a further 23 undergraduates, 25 graduating students, 28 experienced professionals and 10 medical educators. They found that, out of Peterson and Seligman's 24 character strengths, medical students and doctors in the United Kingdom prioritise the virtues of fairness, honesty, judgement, kindness, leadership and teamwork. While medical students and doctors typically rated their own fairness, honesty, kindness and teamwork highly, they were less confident in their own judgement and leadership. Arthur *et al.* also investigated the operation of virtuous, rule-based and consequences-based motivations in how respondents solved common moral dilemmas in medicine and found that, while virtuous motivations had a clear role, depending on the scenario posed, rule-based and consequences-based reasoning were also important. It is important to note that

[12] See Haslam et al. (2004) and Macdonald et al. (2008).

Arthur *et al.*'s study is a survey of attitudes to and views on character amongst medical students and doctors and *not* in itself a psychometric measure to assess the character of individuals. However, with a baseline established as to what respondents' views are regarding these matters, the intention is to use and make available data from this research in order to facilitate the design of a measure of medical ethics that is truly virtue-based rather than principles-based.

Measuring specific virtues in medicine

Next to attempts to measure virtue in general, some researchers have focussed on measuring specific virtues that are relevant to medical practice. While it is not always regarded in virtue-terms, specifically, one of the most studied doctors' virtues is empathy.

Empathy

Defined broadly, empathy can be understood as that personal attribute of the doctor that enables her to understand the patient. (Pedersen, 2008) More specifically, what is called 'clinical empathy' has two dimensions: an affective or emotional dimension that reflects the emotional response that a person shows to what another feels, communicates, experiences, etc. and a cognitive dimension that consists in the ability to understand the emotions, communications, experiences, etc. of another person. (Batt-Rawden *et al.* 2013: 1171) Seen as an amalgam of (at least) emotion and cognition, empathy begins to appear quite like a *virtue*; this makes attempts to study empathy of great relevance for the purposes of this paper.

Hemmerdinger *et al.* (2007) and Pedersen (2009) conducted systematic reviews of instruments to measure empathy and identified 36 and 38 quantitative measures of empathy respectively. Some of the most frequently used instruments are the Interpersonal Reactivity Index (IRI), Jefferson's Scale of Physician Empathy (JSPE) and Hogan's Empathy Scale. Hemmerdinger *et al.* (2007) grouped the 36 measures they identified into three types:

- First person measures (self-reporting measures)
- Second person measures (i.e. patient reports of interactions with doctors)
- Third person measures (observer ratings of empathy in interactions between doctors and patients)

Of these, first person measures are most frequently used.

Important results of research on empathy in medicine include that a decline in empathy occurs during medical school and subsequent hospital-based training (residency). (Neuman *et al.*, 2011) Hojat *et al*. (2002), for instance, found an empathy decline amongst medical students and Bellini and Shea (2005) found a decline in empathy amongst medical trainees as they enter hospital practice. Other researchers have found a correlation between empathy levels and medical specialty and between empathy levels and gender. Hojat *et al.* (2002), for instance, found higher empathy levels amongst psychiatrists than physicians and higher empathy levels amongst women than men. A number of studies also '…indicate a relationship between being pressed for time and lowered empathy…' (Pedersen, 2009: 317)

Importantly, both Pedersen and Hemmerdinger *et al.* point out methodological problems with the existing standardised approaches to measuring empathy, with Hemmerdinger *et al.* concluding that the existing empathy measures are not 'sufficiently reliable and valid for pre-training selection.' (2007: 8) While research on empathy in medicine, then, may serve as a model for research on virtues in medicine, empathy researchers' reliance on self-reporting means that research in this field is not fundamentally more advanced than other attempts to understand virtue psychologically.[13]

Care

Closely related to the virtue of empathy, is the virtue of 'care' or 'caring'. Leffel, Oakes Mueler, Rasinski, Curlin and Yoon (2014) sketch a model of what they call virtuous caring and present data in support of the construct validity of the model. Leffel *et al.* follow Haidt in holding that moral intuitions, emotions and virtues are more explanatory of moral action than ethical reasoning in terms of rules or principles. It follows that the main form of ethical education for doctors should be a form of character education that focuses on the inculcation of these intuitions, emotions and virtues (rather than an education in reasoning in terms of the principles of medical ethics).

Leffel *et al.*'s model of virtuous caring comprises four facets:

- 'care sensitivity' –being perceptually sensitive to signs of suffering and need in others
- 'mindful awareness' – being aware of and attentive to relevant features of situations in which care is appropriate
- 'emotional intensity' –to see and feel the suffering of another and to be motivated to relieve the suffering by some altruistic action
- 'beneficent action' –the tendency to act on the other three components of the model by caring for others when it is appropriate.

In support of their model, Leffel *et al.* investigate (inter alia):

1. Whether the virtue of mindfulness, the emotion-related virtue of emotional intensity/empathic compassion, and the action-related virtue of beneficent action/generosity are positively associated with each other
2. Whether doctors who self-report higher levels of the virtues are also regarded by their peers as virtuous
3. Whether doctors who self-report higher levels of the virtues also experience greater eudaimonic well-being

[13] Notably, efforts have recently been made to understand empathy from a brain-based perspective. Decety *et al.* (2010) performed a study using electrophysiological methods (Event Related Potential or ERP methods) to study physiological reactions to perceiving pain in another. When comparing doctors to a control group, Decety *et al.* found that doctors' physiological response to watching a person's skin pierced by a needle was dampened. What this physiological response alone indicates about the character trait empathy in doctors is debateable; however methods beyond self-reporting are increasingly becoming available.

Leffel *et al.* surveyed 960 medical students from 24 medical schools in the United States. In their questionnaire, they administered eleven measures in total, including measures of mindfulness, empathic compassion, generosity, eudaimonic well-being and meaning in life. Leffel *et al.* found that the mindfulness, empathic compassion and generosity are related. They also found that those who measured higher on generosity and compassion were more likely to be nominated by their peers for the Gold Humanism award (indicating being regarded by their peers as humane). They found that mindfulness, compassion and generosity were correlated significantly with the indices of *eudaimonic* well-being used. Lastly, they found care/harm to be the strongest moral intuition amongst their sample.

Wisdom

Drawing on Aristotle, Pellegrino and Thomasma already held that practical wisdom is an indispensable virtue in medicine in that it coordinates all of the doctor's other (separate) virtues in action. (1993: 84 – 6) More recently, Kaldjian (2010: 560) has suggested that good clinical judgement and practical wisdom are the same thing. This is because the elements of good clinical judgement (the elements of clinical reasoning like forming hypotheses about the cause of symptoms and signs, gathering of information, testing, diagnosis and the recommendation of treatment) appear very much like the elements of wise decisions; both practical wisdom and good clinical judgement must also involve careful and accurate selection of the good for the patient.

While stressing the importance of practical wisdom in medicine, authors such as Pellegrino and Thomasma and Kaldjian do not make specific proposals regarding how to measure or assess wisdom in medicine. In general psychology, however, work on the empirical study of wisdom is growing. Glück *et al.* (2013) survey psychometric measurements of wisdom and find that four measures – the Self-Assessed Wisdom Scale (SAWS), the Three-Dimensional Wisdom Scale (3D-WS), the Adult Self-Transcendence Inventory (ASTI) and the Berlin Wisdom Paradigm (BWP) are already in common use. Glück *et al.* point out, however, that there is no agreed definition of wisdom and that these four measures all measure different aspects or different kinds of wisdom. These kinds of wisdom are:

- 'personal wisdom' – the insights that a person has gained regarding to their own life or experiences and the degree to which they are reflective about their own life (measured by the SAWS, ASTI, aspects of 3D-WS)
- 'general wisdom' – the insights that a person has gained into human life and the human situation generally (measured by the BWP) and
- 'other-related wisdom' – the empathy and concern that a person shows other people or humanity at large (measured by aspects of the 3D-WS).

'Other-related wisdom' is clearly closely associated with 'empathy' (see above). (Glück *et al.*, 2013: 5) The format of the four wisdom measures considered mostly rely on self-reporting (SAWS, ASTI, 3D-WS), although the BWP is a performance-based measure and relies on the participant reporting – in think aloud fashion – how they would respond in a wisdom requiring situation. (Glück *et al.*, 2013: 1)

While the field of wisdom research in psychology is becoming more developed, questions exist regarding the applicability of this research to the medical setting and

its methodological basis. Firstly, as Glück *et al.* point out, most wisdom research in psychology focuses on exploring personal wisdom – that is the insight a person has gained about their own life and experiences. While no doubt an admirable trait in any person, there is no obvious practical link between this personal wisdom and the ability to make good clinical judgements or wise decisions on behalf of patients (that researchers in the medical field would be interested in). Secondly, current wisdom research in psychology relies heavily on self-reporting and in the discussion of general methods to assess virtue in psychology we have already raised some questions about the accuracy of self-reports in establishing virtue in a respondent. Clearly, more research is needed to investigate medical wisdom, specifically and to bring appropriate methods to bear in studying it.

6. Discussion

It should be clear from the above that while there are ample theoretical accounts of medical ethics from a virtue perspective, there is a dearth of empirical work on the virtues (or vices!) or real medical practitioners or on how doctors' virtue develop during programmes of study or during the medical career. In sections 4 and 5, we sketched the criteria that assessments of virtue in medical practice would have to meet and canvassed the – limited – studies from the literature that attempt to measure virtue (or some aspect of it) amongst medical practitioners. It is safe to say that no complete empirical method exists to study the development of morality along virtue lines.

The most general tool that exists for the assessment of virtue is the VIA. However, the VIA suffers from acknowledged problems (see above), including high ceiling effects and the fact that the structure of the character strengths and virtues predicted by the VIA seems faulty (as found by a number of factor analyses of VIA data to date). Moreover, at root the VIA is only a self-report instrument – it reflects nothing more than what respondents say about their own character and self-reports of one's own morality are notoriously susceptible to overestimation, deception and confabulation on the part of the respondent. Studies in the fields of medical ethics or professionalism that have used the VIA (such as Shulz *et al.* 2013) must be judged in this light.

In order to overcome some of the difficulties associated with self-report methodology, it is possible to collect self-report data in different ways than simply asking a person to rate – once and for all – how closely a certain description reflects their character. Jayawickreme *et al.* (2014) describe the utilisation of 'experience sampling methodology' (ESM) to measure virtue. First developed by Larson and Csikszentmihalyi (1983), ESM requires participants to make notes of their personal experiences in the moment over an extended period of time. ESM studies today use smartphones, tablets, etc. to make regular random requests to participants to note something about their current personal experiences. This enables researchers to

> '…observe the frequency with which each individual acts at each level of the personality state dimension, and these frequencies are then combined to form a distribution for each individual of each trait…' (Jayawickreme *et al.* 2014)

In effect, researchers can conduct multiple mini-psychometric tests over an extended period of time and so obtain an average picture of a person's behaviour. Extending self-reporting of behaviour over a period of time and conducting the self-reporting exercises frequently and at random times seems to overcome many of the problems associated with self-reporting.[14]

While ESM extends self-reporting over time and allows for more natural reporting of experience, it still relies on self-reporting. Rather than self-reporting, then, it may be better to rely on methods where others report on a person's virtues than on self-reporting and some such methods have been attempted. Collating reports from peers (for instance parents, friends or spouse) on how virtuous a person is has been the approach in studies by, for instance Hawkins *et al.* (2007) and such peer reports have been found to correlate with self-reports. A growing body of work also exists on the third-personal study of the lives of moral exemplars in depth through biographical methods that may include interviews with the figure being studied as well as interviews with close associates, diary methods, etc. Thus Damon and Colby (1992) have studied the lives of twenty moral exemplars (including Gandhi and Mandela) with an eye to understanding their moral character. And others, e.g. Walker and Frimer (2007) have conducted 'life-review interviews' (in concert with personality questionnaires) in order to establish what it is about the individual personality that predicts moral action. All such methods – in-depth interviews, observations of moral actors, collecting reports from their peers, diary methods, etc. – provide methods whereby *another* person (not the subject herself) could report on the extent to which the subject displays a certain virtue.

Lastly, one could attempt to study how virtuous a person is by experimental methods (i.e. placing subjects in an experimental setting and seeing how they act) (e.g. Decety *et al.*, 2010; Greene and Paxton, 2009) and – even - indirect or biological measures (such as brain scanning, measuring reaction times or measuring involuntary reactions). However, it is safe to say that such methods are in their infancy.

More fine-grained than general assessments of virtue have been the attempts to measure specific virtues in medicine, however, here, too, methods are in their infancy. The virtue most studied in the medical field appears to be empathy and a large number of measures of empathy exist. Studies on a considerable scale (such as Leffel *et al.*'s study) are also now being conducted in the area. Lastly, psychometrically validated wisdom measures also exist, but have as yet not found traction in the study of medical ethics.

7. Conclusion

Virtue-based approaches to medical ethics and professionalism seem to hold great promise. Virtue ethics holds that the moral character of the doctor is of the greatest importance in assessing whether their practice of medicine is ethical and professional. It follows that assessment of the ethicality and professionalism with

[14] For more on the use of ESM in this area, see Fleeson (2007) and Fowers & Lefevor (2013). For one study using ESM in medicine (although not to study character specifically), see Ahmad *et al.* (2012)

which doctors practice will have to proceed principally by assessment of doctors' own moral character.

We surveyed the literature and found that most work on assessing ethics and professionalism in medicine is conducted from a perspective that is rationalistic or cognitivistic (and is rules- or principles-based) and not virtue-based. We also found that, of those few studies that are virtue-based or that lend themselves to a virtue-interpretation, none yet lives up to the billing of a true virtue-based assessment of doctors' ethics.

Until researchers who take a virtue approach to medical ethics can begin to operationalise and study virtue concepts in medical education and the medical career, it will be hard to turn virtue-insights into teaching interventions or regulation programmes that are demonstrably effective. True measures of doctors' virtue would be capable of testing for respondents' moral sensitivity, emotion and reasoning all at once and would have to be able to assess the extent to which these factors are in alignment. They would seek to build on the work already done to measure specific virtues relevant to medicine such as empathy and wisdom. However, researchers should be wary of relying on simple self-reporting and need to adopt some of the measures that mitigate the problems to do with self-reporting. Instead of self-reporting, the use of Experience Sampling Methodology (ESM) promises much. Alternatively approaches that seek to gain second or third personal views on respondents' virtue development may also provide answers in this challenging area.

References

Akabayashi, A., Slingsby, B., Kai, I., Nishimura, T. and Yamagishi, A. 2004. 'The Development of a Brief and Objective Method for Evaluating Moral Sensitivity and Reasoning in Medical Students' *BMC Medical Ethics* 5 (1) doi:10.1186/1472-6939-5-1

Ahmad, A., Schneider, J., Nwadei, I., Darcy, M., Farr, R., Thal, R. and Arora, V. 2012. 'A Pilot Study of Real-Time Experience Sampling Method to Evaluate Student Engagement in a Global Health Rotation' *Education Research International* http://dx.doi.org/10.1155/2012/241874

Annas, J. 2011 *Intelligent Virtue*, Oxford: Oxford University Press.

Anscombe, E. 1958 'Modern Moral Philosophy', *Philosophy,* vol. 33, no. 124, pp. 1-19.

Aristotle (1999) *Nichomachean Ethics,* trans. T. Irwin, Indianapolis: Hackett.

Arthur, J., Kristjansson, K., Thomas, H., Kotzee, B., Ignatowicz, A. and Qiu, T. 2015. *Virtuous Medical Practice*. Birmingham: University of Birmingham.

Baldwin, D. and Self, D. (2005) 'The Assessment of Moral Reasoning and Professionalism in Medical Education and Practice', in Stern, D. (ed.) *Measuring Medical Professionalism*, Oxford: Oxford University Press.

Barilan, Y. and Brusa, M. (2012) 'Deliberation at the Hub of Medical Education: Beyond Virtue Ethics and Codes of Practice', *Medicine, Health Care and Philosophy,* [Online], Available at: http://www.ncbi.nlm.nih.gov/pubmed/22740074 [Accessed: 13 November 2014].

Batt-Rawden, S., Chisolm, M., Anton, B., Flickinger, T. 2013. 'Teaching Empathy to Medical Students: an updated, systematic review. *Academic Medicine* 88 (8), 1171 – 7.

Beauchamp, T. and Childress, J. (1979) *Principles of Biomedical Ethics*, Oxford: Oxford University Press.

Bebeau, M. (2002) 'The Defining Issues Test and the Four Component Model: Contributions to Professional Education', *Journal of Moral Education,* vol. 31, no. 3, pp. 271-95.

Bebeau, M. (2006) 'Evidence-Based Character Development', in: Kenny, N. and Shelton, W. *Lost Virtue: Professional Character Development in Medical Education: Advances in Bioethics, vol. 10*, Oxford: JAI Press.

Bebeau, M., Born, D. and Ozar, D. 1993. 'The Development of a Professional Role Orientation Inventory' *Journal of the American College of Dentists* 60 (2), 27 – 33.

Bebeau, M., Rest, J. and Yamoor, C. 1985. 'Measuring Dental Students' Ethical Sensitivity' *Journal of Dental Education* 49 (4), 225 – 35.

Bebeau, M. and Thoma S. 1999. '"Intermediate" concepts and the connection to moral education' *Educational Psychology Review* 11, 343–360.

Bellini, L. and Shea, J. 2005. 'Mood change and empathy decline persist during three years of internal medicine training' *Academic Medicine* 80 (2), 164 – 7.

Bergman, M., Drasgow, F., Donovan, M., Henning, J. and Juraska, S (2006) 'Scoring Situational Judgment Tests: Once You Get the Data, Your Troubles Begin', *International Journal of Selection and Assessment,* vol. 14, no. 3, pp. 223-35.

Bertua, C., Anderson, N., and Salgado, J. (2005) 'The Predictive Validity of Cognitive Ability Tests: A UK Meta-analysis', *Journal of Occupational and Organisational Psychology*, vol. 78, no. 3, pp. 387-409.

Blasi, A. 1980. 'Bridging Moral Cognition and Moral Action: a review of the literature' *Psychological Bulletin* 88 (1), 1 – 45.

Borgstrom, E., Cohn, S. and Barclay, S. (2010) 'Medical Professionalism: Conflicting Values for Tomorrow's Doctors', *Journal of General Internal Medicine,* vol. 25, no. 12, pp.1330-1336.

Born, D., Bebeau, M. and Rozmenoski, S. 1995. 'Analysis of the Professional Role Orientation Inventory' *Journal of Dental Research* 74, 162 – 62.

Braun, U., Gill, A., Teal, C. and Morrison, L. 2013. 'The Utility of Reflective Writing after a Palliative Care Experience: can we assess medical students' professionalism? *Journal of Palliative Medicine* 16 (11), 342 – 9.

Bryan, C. and Babelay, A. (2009) 'Building Character: A Model for Reflective Practice', *Academic Medicine*, vol. 84, no. 9, pp. 1283-1288.

Caldicott, C., Faber-Langendoen, K., Bebeau, M. and Thoma, S. (2010) 'Assessing Moral Reasoning in Medical Training and Practice: A Pilot Study', *Cambridge Consortium for Bioethics Education*, New York City, USA.

Calman, K. and Downey, R. (1987) 'Practical Problems in the Teaching of Ethics to Medical Students', *Journal of Medical Ethics*, vol. 13, pp. 153-156.

Campbell, A., Chin, J. and Voo, T-C. (2007) 'How Can We Know that Ethics Education Produces Ethical Doctors?', *Medical Teacher,* vol. 29, no. 5, pp. 431-436.

Carpendale, J. 2000. 'Kohlberg and Piaget on Stages and Moral Reasoning' *Developmental Review* 20 (2), 181 – 205.

Cawley, M., Martin, J. and Johnson, J. 2000. 'A Virtues Approach to Personality' *Personality and Individual Differences* 28 (5), 997 – 1013.

Cowley, C. (2005) 'The Dangers of Medical Ethics', *Journal of Medical Ethics*, vol. 31, no. 12, pp. 739-742.

Coulehan, J. (2005) 'Today's Professionalism: Engaging the Mind But not the Heart', *Academic Medicine,* vol. 80, no. 10, pp. 892-898.

Curren, R. and Kotzee, B. 2014. 'Can Virtue Be Measured?' *Theory and Research in Education* vol. 12, no. 3, pp. 266 – 82.

Damon, W. and Colby, A. 1992. *Some Do Care: contemporary lives of moral commitment*. New York: Free Press.

Decety, J., Yang, C. and Cheng, Y. 2010. 'Physicians Down-Regulate their Pain Empathy Response: an event-related brain potential study' *NeuroImage* 50, 1676 – 82.

Doris, J. 2002. *Lack of Character: Personality and Moral Behaviour.* Cambridge: CUP.

Drane, J. F. (1995) *Becoming a Good Doctor: The Place of Virtues and Character in Medical Ethics*, 2nd edition, Kansas: Sheed and Ward.

Eckles, R., Meslin, E., Gaffney, M, and Helft, P. (2005) 'Medical Ethics Education: Where Are We? Where Should We Be Going? A review', *Academic Medicine*, vol. 80, no. 12, pp. 1143-1152.

Fowers, B. 2005. *Virtues and Psychology: pursuing excellence in ordinary practices*. Washington, DC: APA Press.

Fowers, B. (2014) 'Toward Programmatic Research on Virtue Assessment: Challenges and Prospects', *Theory and Research in Education*, vol. 12, no. 3, pp. 309-328.

Goldie, J.G., Dowie, A., Cotton, P., and Morrison, J.M. 2007. *Teaching professionalism in the early years of a medical curriculum: a qualitative study. Medical Education* 41 (6), 610-617.

Goldie, J. (2012) 'The formation of professional identity in medical students: Considerations for educators', *Medical Teacher,* vol. 34, no. 9, pp. 641-648.

Greene, J. and Paxton J. 2009. 'Patterns of neural activity associated with honest and dishonest moral decisions' *Proceedings of the National Academy of Sciences of the United States of America* 106: 12506–12511.

Haidt, J. 2001. 'The Emotional Dog and its Rational Tail: a social intuitionist approach to moral judgement' *Psychological Review* 108 (4), 814 – 34.

Haidt, J. 2012. *The Righteous Mind: why good people are divided by politics and religion*. London: Penguin.

Harkness, S., Edwards, C. and Super, C. 1981. 'The Claim to Moral Adequacy of a Highest Stage of Moral Judgement' *Developmental Psychology* 17 (5), 595 – 603.

Harman, G. 1999. 'Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error'. *Proceedings of the Aristotelian Society* 99, 315 – 31.

Haslam, N., Bain, P., and Neal, D. 2004. 'The Implicit Structure of Positive Characteristics' *Personality and Social Psychology Bulletin* 30, 529–541.

Hawkins, A., Fowers, B., Carroll, J. and Yang, C. 2007. 'Conceptualising and Measuring Marital Virtues'. In: Hofferth, S. and Casper, L. (eds) *Handbook of Measurement Issues in Family Research*. Hillsdale, NJ: Lawrence Erlbaum.

Hebert, P., Meslin, E., Dunn, E., Byrne, N. and Reid, S. 1990. 'Evaluating Ethical Sensitivity in Medical Students: using vignettes as instrument' *Journal of Medical Ethics* 16 (3) 141 – 5.

Hebert, P., Meslin, E. and Dunn, E. 1992. 'Measuring the Ethical Sensitivity of Medical Students: a study at the University of Toronto' *Journal of Medical Ethics* 18 (3), 142 – 7.

Helkama, K., Uutela, A., Pohjanheimo, E., Salminen, S., Koponen, A., Rantanen, V. and Ntsi, L. 2003. 'Moral Reasoning and Values in Medical School: a longitudinal study in Finland' *Scandinavian Journal of Educational Research* 47 (4), 399 – 411.

Hemmerdinger, J., Stoddart, S. and Lilford, R. 2007. 'A Systematic Review of Tests of Empathy in Medicine' *BMC Medical Education* 7: 24

Hojat, M., Gonnella, J., Mangioni, S., Naska, T., Veloski, J., Erdman, J., Callahan, C. and Magee, M. Empathy in Medical students as related to academic performance, clinical competence and gender. *Medical Education* 36 (6), 522 – 7.

Hren, D., Marusic, M and Marusic, A. 2011. 'Regression of Moral Reasoning during Medical Education: Combined Design Study to Evaluate the Effect of Clinical Study Years' *PLOS One* DOI: 10.1371/journal.pone.0017406

Hursthouse, R. 1999. *On Virtue Ethics*. Oxford: OUP.

Hursthouse, R. 2011. 'What Does the Aristotelian *Phronimos* Know?' In: Jost, L. and Wuerth, J. (eds.) *Perfecting Virtue: New Essays on Kantian Ethics and Virtue Ethics*. Cambridge: CUP.

Jayawickreme, E., Meindl, P., Helzer, E., Furr, M. and Fleeson, W. 2014. 'Virtuous States and Virtuous Traits: how the empirical evidence regarding the existence of broad traits saves virtue ethics from the situationist critique. *Theory and Research in Education* 12 (3), 283 – 308.

Jha, V., Bekker, H. L. and Duffy, S. R. G. (2007) 'A Systematic Review of Studies Assessing and Facilitating Attitudes Towards Professionalism in Medicine', *Medical Education,* vol. 41, no. 8, pp. 822-829.

Kaldjian, L. C. (2010) 'Teaching Practical Wisdom in Medicine Through Clinical Judgement, Goals of Care, and Ethical Reasoning', *Journal of Medical Ethics,* vol. 36, no. 9, pp. 558-562.

Kamtekar, R. 2004. 'Situationism and Virtue Ethics on the Content of Our Character' *Ethics* 114, 458 – 91.

Kinghorn, W. 2010. 'Medical Education as Moral Formation: an Aristotelian account of medical professionalism' *Perspectives in Biology and Medicine* 53 (1), 87 – 105.

Kipnes, K. and Gerhard, A. 1995. 'Some Ethical Principles for Adult Critical Care' *Hawaii Medical Journal* 54 (4), 482 – 5.

Kohlberg, L. 1981. *The Philosophy of Moral Development.* San Francisco: Harper and Row.

Kohlberg, L. 1984. *The Psychology of Moral Development.* San Francisco: Harper and Row.

Kristjánsson, K. (2015) *Aristotelian Character Education,* London: Routledge.

Lantz, M., Bebeau, M. and Zarkowski, P. 2011. 'The Status of Teaching and Learning of Ethics in US Dental Schools'.*Journal of Dental Education*, 75 (10) 1295-1309.

Larkin, G., Iserson, K., Kssutto, Z., Freas, G., Delaney, K., Krimm, J., Schmidt, T., Simon, J., Calkins, A. and Adams, J. (2009) 'Virtue in Emergency Medicine', *Academic Emergency Medicine*, vol. 16, no. 1, pp. 51 – 55.

Larson, R. and Csikszentmihalyi, M. 1983. 'The Experience Sampling Method' *New Directions for Methodology of Social and Behavioral Science* 15, 41 – 56.

Leffel, G., Oakes Mueller, R., Curlin, F. and Yoon, J. 2014. 'Relevance of the Rationalist-Intuitionist Debate for Ethics and Professionalism in Medical Education' *Advances in Health Sciences Education.*

Lohfeld, L., Norman, G., Goldie, J., Schwartz, L., Eva, K., Cotton, P., Morrison, J., Kulmakan, K. and Wood, T. 2012. 'Testing the Reliability and Validity of a Scenario-Based Questionnaire to Assess the Ethical Sensitivity of Undergraduate Medical Students' *Medical Teacher* 34 (8), 635 – 42.

Macdonald, C., Bore, M., and Munro, D. 2008. 'Values in Action scale and the big 5: an empirical indication of structure. *Journal of Research in Personality* 42, 787-799.

MacIntyre, A. (1984) *After Virtue*, Notre Dame: University of Notre Dame Press.

MacIntyre, A. (1990) *Three Rival Versions of Moral Inquiry*, Notre Dame: University of Notre Dame Press.

Mattick, K. and Bligh, J. (2006) 'Teaching and Assessing Medical Ethics: Where Are We Now?', *Journal of Medical Ethics,* vol. 32, no. 3, pp. 181-185.

Neuman, M., Edelhauser, F., Tauschel, D., Fisher, M., Wirtz, M., Woopen, C., Haramati, A. and Scheffer, C. 2011. 'Empathy Decline and Its Reasons: a systematic review of studies with medical students and residents' *Academic Medicine* 86 (8), 996 – 1009.

Noftle, E., Schnitker, S. and Robins, R. 2011. 'Character and Personality: connections between positive psychology and personality psychology' In: Sheldon, K., Kashdan, T and Steger, M. (eds) *Designing Positive Psychology: taking stock and moving forward.* Oxford: Oxford University Press.

Oakley, J. and Cocking, D. 2001. *Virtue Ethics and Professional Roles*. Cambridge: Cambridge University Press.

Patenaude, J., Niyonsenga, T. and Faffard, D. 2003. 'Changes in students' moral development during medical school: a cohort study' *CMAJ* 168 (7), 840 – 4.

Pedersen, R. 2008. 'Empathy: a wolf in sheep's clothing?' *Medicine, Health Care and Philosophy* 11 (3), 325 – 35.

Pedersen, R. 2009. 'Empirical Research on Empathy in Medicine – a critical review' *Patient Education and Counselling* 76 (3), 307 – 22.

Pellegrino, E., & Thomasma, D. (1993) *The Virtues in Clinical Practice,* New York: Oxford University Press.

Peterson, C. and Seligman, M. (2004) *Character Strengths and Virtues: A Handbook and Classification*, Oxford: Oxford University Press.

Pinijphon, P. (2009) *An ICM Approach to the Assessment of a Medical Ethics Intervention in Thailand,* Ph.D., University of Alabama.

Ramakrishna, J., Valani, R., Sriharan, A. and Scolnik, D. 2014. 'Design and Pilot Implementation of an Evaluation Tool Assessing Professionalism, Communication and Collaboration During a Unique Global Health Elective' *Medicine Conflict and Survival* 30 (1), 56 – 65.

Rest, J.R. (1979) *Development in Judging Moral Issues*, USA: University of Minnesota Press.

Rhodes, R., and D. S. Cohen (2003) 'Understanding, Being and Doing: Medical Ethics in Medical Education', *Cambridge Quarterly of Healthcare Ethics*, vol. 12, no. 1, pp. 39-53.

Sanderse, W. 2012. *Character Education: A Neo-Aristotelian Approach to the Philosophy, Psychology and Education of Virtue.* Delft: Eburon Uitgeverij.

Schmidt, H.G., Norman, G.R. and Boshuizen, H.P.A. (1990) 'A Cognitive Perspective on Medical Expertise: Theory and Implications', *Academic Medicine*, vol. 65, no. 10, pp. 611-621.

Schulz, K., Puscas, L., Tucci, D., Woodard, C., Witsell, D., Esclamado, R. and Lee, W. (2013) 'Surgical Training and Education in Promoting Professionalism: A Comparative Assessment of Virtue-based Leadership Development in Otolaryngology-head and Neck Surgery Residents', *Medical Education Online,* vol. 18, [Online], Available at: http://med-ed-online.net/index.php/meo/article/view/22440 [Accessed: 13 November 2014].

Schwartz, B. and Sharpe, K. (2010) *Practical Wisdom: the right way to do the right thing*, New York: Riverhead Books.

Self, D. and Baldwin, D. (1994) 'Moral Reasoning in Medicine', in Rest, J., Narvaez, D. (eds.) *Moral Development in the Professions*, Hillsdale (NJ): Lawrence Erlbaum.

Smith, S., Balint, J., Krause, K., Moore-West, M. and Viles, P. 1994. 'Performance-Based Assessment of Moral Reasoning and Ethical Judgement Among Medical Students' *Academic Medicine* 69 (5), 381 – 6.

Snow, N. (2010) *Virtue as Social Intelligence: an empirically grounded theory*, London: Routledge.

Sreenivasan, G. 2002. 'Errors about Errors: Virtue Theory and Trait Attribution' *Mind* 111 (441), 47 – 68.

Stirrat, G. M., Johnston, C., Gillon, R., & Boyd, K. (2010) 'Medical Ethics and Law for Doctors of Tomorrow: The 1998 Consensus Statement Updated', *Journal of Medical Ethics,* vol. 36, no. 1, pp. 55-60.

Tallis, R. (2006) 'Doctors in Society: Medical Professionalism in a Changing World', *Clinical Medicine,* vol. 6, no. 1, pp. 7-12.

Thoma, S., Bebeau, M. and Born, D. 1998. 'Further Analysis of the Professional Role Orientation Inventory' *Journal of Dental Research* 77, 116 – 20.

Toon, P. (2014) *A Flourishing Practice?*, London: Royal College of General Practitioners.

Tsai, T., Harasym, P.H., Coderre, S., Mclaughin, K. and Donnon, T. 2009. Assessing ethical problem solving by reasoning rather than decision making. *Medical Education* 43 (12), 1188-1197.

Veatch, R. M. (2006) 'Character Formation in Professional Education: A Word of Caution', in Kenny, N, and Shelton, W. (eds.) *Lost Virtue: Professional Character Development in Medical Ethics*, Amsterdam: Elsevior.

Walker, L. and Frimer, J. 2007. 'Moral Personality of Brave and Caring Exemplars' *Journal of Personality and Social Psychology* 93 (5), 845 – 60.

Wilkinson, T., Wade, W. and Knock, L. 2009. 'A Blueprint to Assess Professionalism' *Academic Medicine* 84 (5), 551 – 8.